Estadística para los no estadísticos

Epidemiología Clínica. Inferencia causal. Error aleatorio.

- **Deriva del hecho de tomar sólo una muestra de la población teórica sobre la que queremos sacar conclusiones.**
- **Su importancia puede disminuirse aumentando el tamaño de la muestra.**
- **Su importancia puede cuantificarse mediante test de hipótesis (probabilidad de error al rechazar la hipótesis nula de igualdad) y/o el cálculo de los intervalos de confianza (valor del estimador muestral con rango poblacional).**
- **La ausencia de error aleatorio se denomina precisión.**

Error aleatorio. Introducción a favor.

- **“Y así como el analfabetismo cierra a quien lo padece casi todas las puertas de la cultura, el anaritmetismo impide el acceso a esa puerta maestra del conocimiento objetivo que es la ciencia. Por supuesto, el discurso científico también utiliza el lenguaje verbal, incluso más que el numérico; pero es en la cuantificación y en la correlación matemática de las magnitudes donde la ciencia se realiza como tal, donde se vuelve concreta, precisa y eficaz.**

Carlo Frabetti. Anaritmetismo. El Pais, miércoles 24 de enero de 1996.

Error aleatorio. Introducción en contra.

- **“La estadística es una disciplina que debe una gran parte de su configuración actual a las matemáticas. Las matemáticas, por otro lado, son un conjunto de conocimientos teóricos, abstractos, cuyos creadores, la mayoría de la veces, no pretendían resolver problemas prácticos, sino generar abstracciones a partir de los números, deducir otros nuevos conocimientos teóricos, derivables de abstracciones previas o resolver problemas, incoherencias o curiosidades generadas, como efectos secundarios, en la creación de teorías matemáticas. Considerar los conocimientos matemáticos como el resultado de mentes creadoras que jugaron con símbolos o aceptaron retos de otros jugadores similares, no es alejarse excesivamente de la historia de esta ciencia.**

Manzano V. Inferencia estadística. Aplicaciones con SPSS/PC+. Madrid: RA-MA, 1995; 98.

Estadística. Proceso de medición.

- **La Estadística es la disciplina encargada del tratamiento de los datos numéricos derivados de los grupos de indivíduos.**
- **La Medición es el proceso por el que se representan las propiedades de los fenómenos investigados mediante números o nombres.**
- **Unidades de análisis son las entidades objeto de nuestro estudio (individuos, grupos, ciudades, hospitales, etc).**
- **Las unidades de análisis tienen características, que varian (variables) y son objeto de comparación entre ellas.**

Estadística. Proceso de medición.

- En cada una de estas características diferentes se incluyen dimensiones o modalidades (niveles, categorias, atributos, valores).
- Entre las modalidades hay relaciones, en ocasiones escasas y simples y otras veces múltiples y complejas.
- Las Escalas de Medida de estas modalidades establecen dos tipos de variables: Cualitativas y Cuantitativas.

Estadística. Escalas de medida.

Tipos de Escalas	Variables	Relación entre valores	Ejemplos
Nominal	Nominales o Categóricas (Di o Policotómicas	Igualdad/desigualdad	Raza, sexo, religión, estado civil, profesión
Ordinal	Ordinales	Igualdad/desigualdad Orden	Nivel socioeconómico, grado de mejoría o empeoramiento, opiniones
De Intervalo	Cuantitivativas contínuas o discretas	Igualdad/desigualdad Orden y Unidad de medida empírica	Grados centígrados, tiempo o longitud sin origen determinado
De Razón	Cuantitativas contínuas o Discretas	Igualdad/desigualdad Orden y unidad empírica de medida, con cero absoluto	Tensión arterial, edad, peso, grados Kelvin.

Las variables de las escalas nominal y ordinal son cualitativas.

Estadística. Diseño de un cuestionario y transferencia a soporte en disco. Algunos puntos destacables.

- **Suele tenerse la tentación de recoger más información de la que realmente se necesita, en previsión de que pueda usarse en el futuro. Recuerda, si sólo tienes que rellenar 20 items/indivíduo, les dedicarás más atención que si rellenas 50.**
- **Define bien las variables.**
- **Para rellenar la base de datos estamos precisamente en este curso; y casi, casi, sólo hacemos el curso para ésto.**
- **Guarda alguna copia de seguridad. ¡Si no lo haces, te acordarás de esta advertencia!.**
- **En general, la matriz de datos muestra la información en filas (pacientes) y columnas (variables).**

Estadística descriptiva.

- Resume los valores que toman las variables en las unidades de análisis.
- Atención al examen de los datos, previo a la realización de técnicas más complejas de confirmación de hipótesis.
- Importancia central de la representación gráfica.
- Resistencia de los estadísticos a valores extremos.
- Distinción entre ajuste y resíduo.
- Apertura a la transformación de variables para conseguir modelos más ajustados.

Estadística descriptiva y Estadística inferencial.

- **Estadística descriptiva.**
  - **Resumen de los datos para condensar la información.**
- **Estadística inferencial.**
  - **Inferir los valores de la población (parámetros) basándonos en el conocimiento de los valores de la muestra (estadísticos).**

Estadística descriptiva.

- **Resumen de los valores que toman las variables en las unidades de análisis:**

- **Variables cualitativas:**
  - **Tablas de distribución de frecuencias absolutas o relativas.**
  - **Gráficas (diagramas de barras y tartas).**

- **Variables cuantitativas:**
  - **Medidas de tendencia central, de variabilidad, de posición, de simetría y de apuntamiento.**
  - **Gráficas (histogramas, polígonos de frecuencias, arbol y hoja, caja y bigotes, nube de puntos, etc.).**

Estadística descriptiva. Variables cualitativas

- **Tablas de distribución de frecuencias absolutas o relativas.**
  - **Se disponen en filas los valores de la variable y en columnas las frecuencias absolutas, relativas y acumuladas.**
  - **Las variables cuantitativas pueden agrupar los valores en intervalos (categorización) y representarse también en una tabla de distribución de frecuencias o categorizarse.**

Estadística descriptiva. Variables cuantitativas

- **Variables cuantitativas:**
  - **Medidas de tendencia central.**
    - **Media aritmética:** $\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n}$

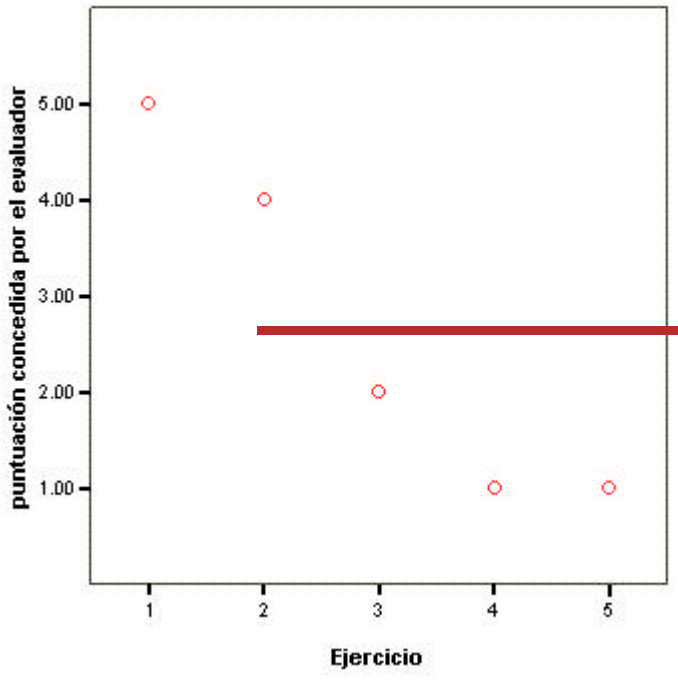
      (Distribuciones normales)
    - **Mediana: Valor que deja 50% de los casos a ambos lados. Igual que el $P_{50.}$ (distribuciones que no son normales).**
    - **Moda: Valor más frecuente.**

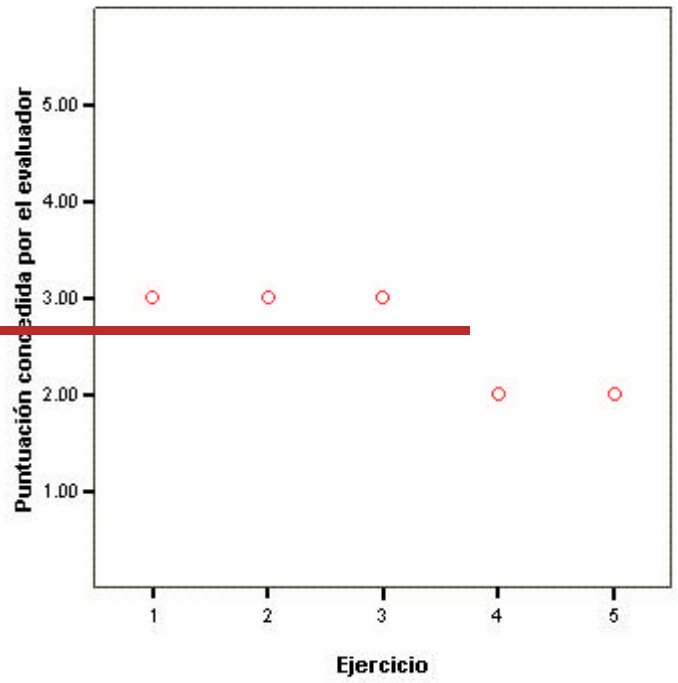
Estadística descriptiva. Variables cuantitativas

- **Variables cuantitativas:**
  - **Medidas de de variabilidad (dispersión).**
    - **Rango: Distancia entre mayor y menor.**
    - **Percentiles. Valor bajo el que se encuentra una cierta proporción (cuartiles, deciles, n-tiles).**
    - **Recorrido intercuartílico: P_{25} a P_{75}.**
    - **Varianza:** $s^2 = \frac{\sum_1^n \left(x_i - \bar{x}\right)^2}{n-1}$
    - **Desviación estandar:** $s = \sqrt{\frac{\sum_1^n \left(x_i - \bar{x}\right)^2}{n-1}}$
    - **Coeficiente de variación:** $CV(x) = \frac{s}{\bar{x}} 100\%$
    - **Coeficiente Variación intercuartílico:** $CVI = \frac{Q_1 - Q_3}{Q_1 + Q_3}$

Ajuste de un modelo

- **Media:** Modelo para resumir nuestros datos.
- **Varianza:** Error promediado para evaluar el ajuste del modelo a los valores de los sujetos.
- Como **Desviación Estandar** se expresa en las mismas unidades que la media.





$$\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n}$$

$$s^2 = \frac{\sum_{1}^{n} \left(x_i - \overline{x}\right)^2}{n-1}$$

$$s = \sqrt{\frac{\sum_{1}^{n} \left(x_i - \overline{x}\right)^2}{n-1}}$$

Estadística descriptiva. Variables cualitativas

- **Gráficas (diagramas de barras y tartas o sectores).**
  - **Utilidades:**
    - **Presentar la información.**
    - **Evaluar la estructura de los datos.**
  - **Tipos:**
    - **Comparaciones de dos o más números (diagrama de barras o pictogramas).**
    - **Distribución de objetos individuales o medidas en diferentes categorías (diagrama de sectores).**
    - **Mostrar el cambio en alguna cantidad con el paso del tiempo (diagrama de líneas)**
    - **Mostrar la relación entre dos mediciones (nube de puntos).**

Estadística descriptiva. Variables cuantitativas

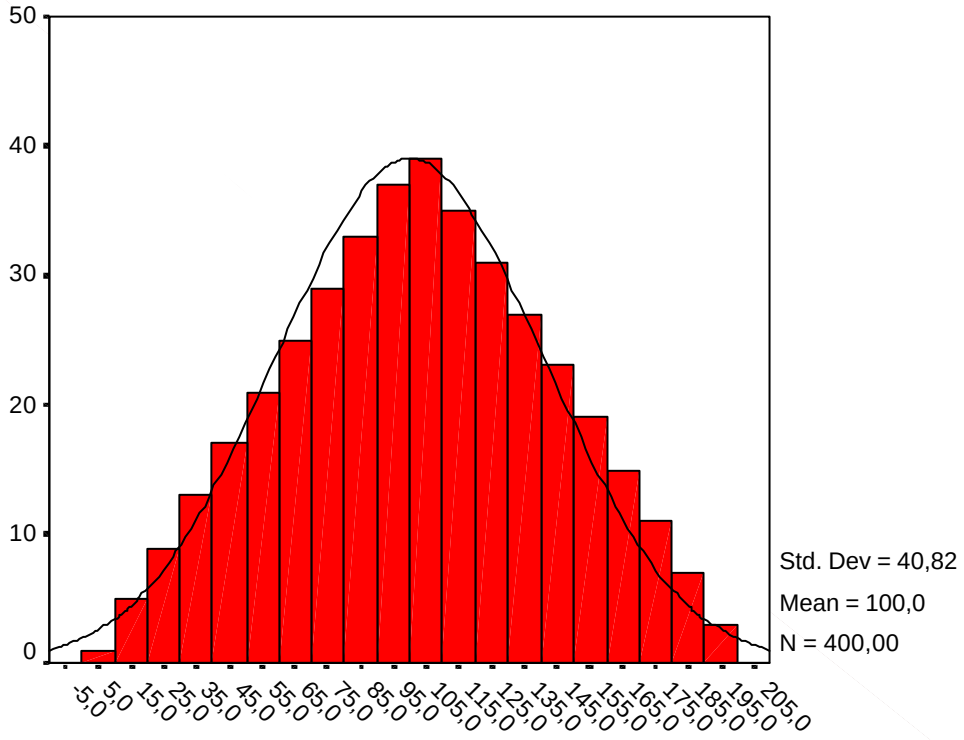
- **Variables cuantitativas:**
  - **Gráficas (histogramas y polígonos de frecuencias).**
    - **Intervalos de clase, marca de clase, rango del intervalo. Pedir 22 intervalos para este fichero.**

Estadística descriptiva.
Variables cuantitativas

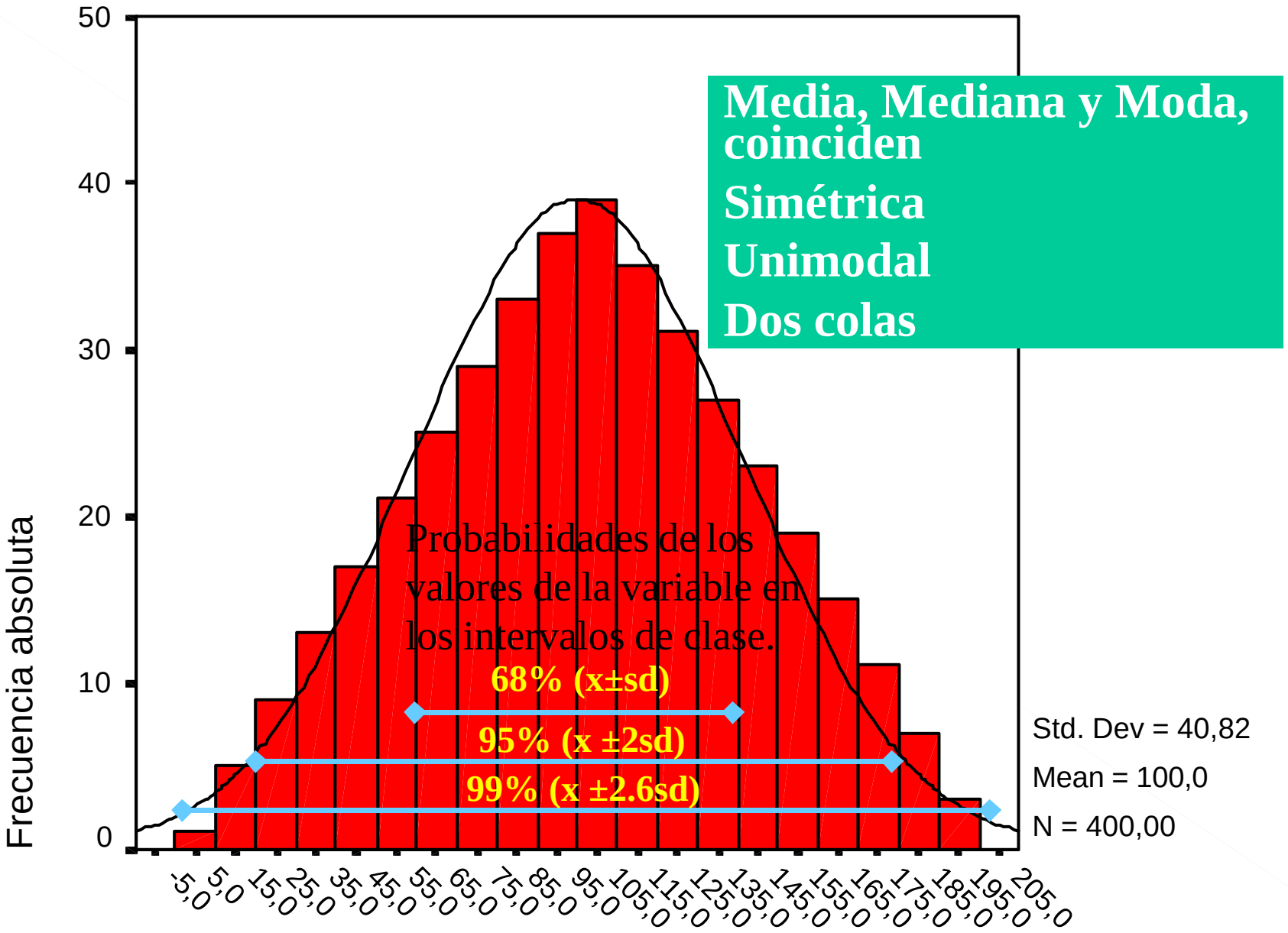
La distribución Normal

Satisfacción del usuario en puntuación

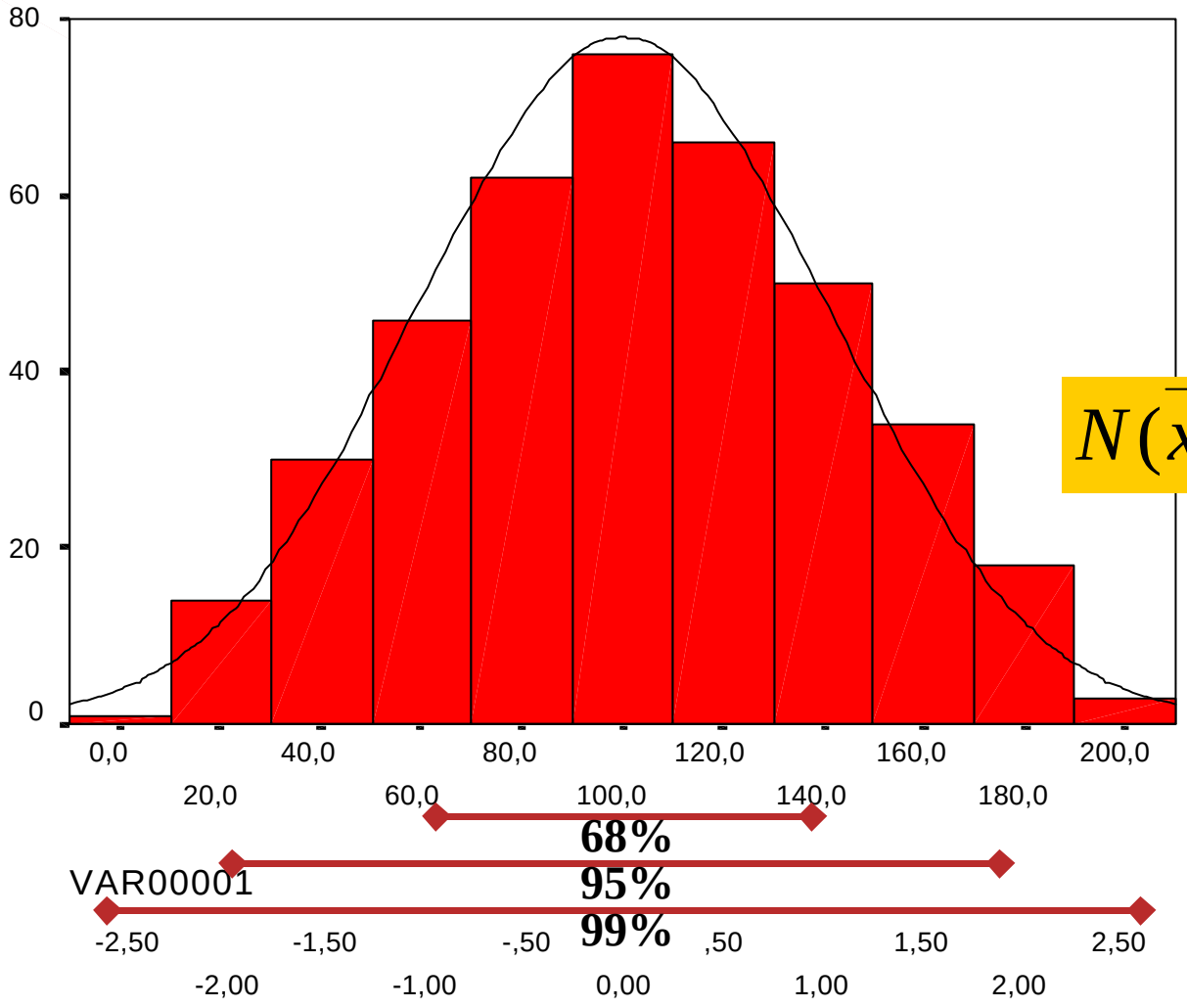
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	5,00	1	,3	,3	,3
	10,00	2	,5	,5	,8
	15,00	3	,8	,8	1,5
	20,00	4	1,0	1,0	2,5
	25,00	5	1,3	1,3	3,8
	30,00	6	1,5	1,5	5,3
	35,00	7	1,8	1,8	7,0
	40,00	8	2,0	2,0	9,0
	45,00	9	2,3	2,3	11,3
	50,00	10	2,5	2,5	13,8
	55,00	11	2,8	2,8	16,5
	60,00	12	3,0	3,0	19,5
	65,00	13	3,3	3,3	22,8
	70,00	14	3,5	3,5	26,3
	75,00	15	3,8	3,8	30,0
	80,00	16	4,0	4,0	34,0
	85,00	17	4,3	4,3	38,3
	90,00	18	4,5	4,5	42,8
	95,00	19	4,8	4,8	47,5
	100,00	20	5,0	5,0	52,5
	105,00	19	4,8	4,8	57,3
	110,00	18	4,5	4,5	61,8
	115,00	17	4,3	4,3	66,0
	120,00	16	4,0	4,0	70,0
	125,00	15	3,8	3,8	73,8
	130,00	14	3,5	3,5	77,3
	135,00	13	3,3	3,3	80,5
	140,00	12	3,0	3,0	83,5
	145,00	11	2,8	2,8	86,3
	150,00	10	2,5	2,5	88,8
	155,00	9	2,3	2,3	91,0
	160,00	8	2,0	2,0	93,0
	165,00	7	1,8	1,8	94,8
	170,00	6	1,5	1,5	96,3
	175,00	5	1,3	1,3	97,5
	180,00	4	1,0	1,0	98,5
	185,00	3	,8	,8	99,3
	190,00	2	,5	,5	99,8
	195,00	1	,3	,3	100,0
	Total	400	100,0	100,0	



La distribución normal



Estadística descriptiva. Estandarización de los valores de una variable con distribución normal.



$$Z = \frac{x_i - \bar{x}}{s}$$

$$N(\bar{x}, s^2) \rightarrow N(0,1)$$

Std. Dev = 40,82

Mean = 100,0

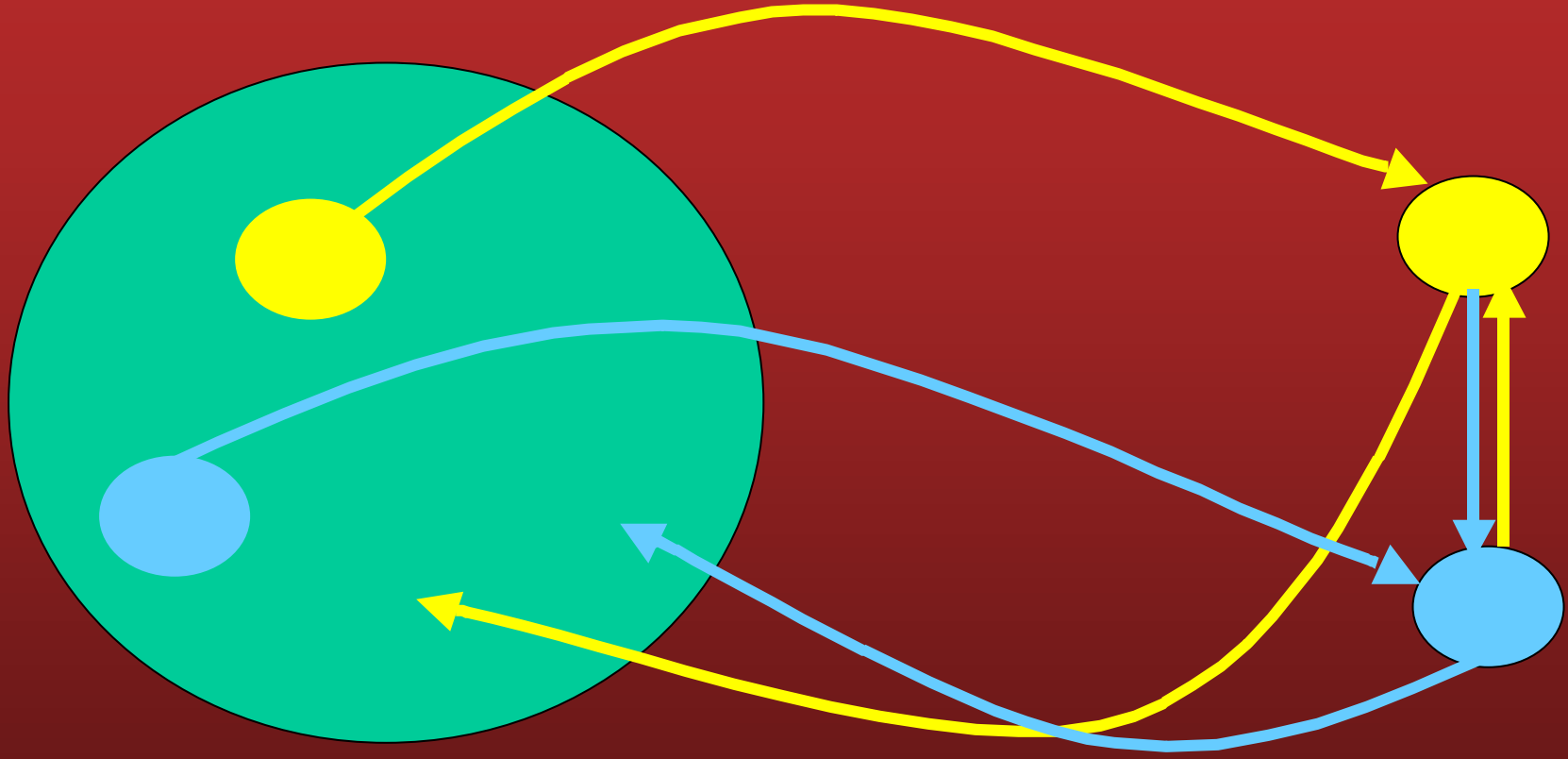
N = 400,00

Estadística descriptiva. Trasformación de variables

Tipo de transformación	Fórmula	Cambios en forma de distribución	Ejemplo
Aritmética			
Lineal	$x_i' = a + bx_i$	Ausentes	Tipificación
Potencia	$x_i' = x_i^q$ $x_i' = \log x$ $x_i' = -x_i^q$	Reducción de asimetría. Si q<1 a la izq. si q> 1 a la dcha. Cambio de curtosis	
Logaritmos y raíces plegadas		Más representativos los valores centrales.	
Lógica			

Estadística inferencial.

Muestreo aleatorio



Estadística Inferencial: Estimación de un parámetro poblacional y pertenencia de dos muestras a la misma población.

Intervalo de confianza y Test de Hipótesis

Estadística inferencial.

- **Muestreo: Permite aplicar la estadística inferencial.**
- **Estadística Inferencial, permite hacer inferencias sobre los valores poblacionales:**
  - **Estima los parámetros poblacionales conociendo los estadísticos muestrales.**
  - **Evalua si dos o más muestras pertenecen a la misma población.**
- **Métodos de inferencia:**
  - **Intervalo de confianza.**
  - **Test de Hipótesis.**

Intervalo de confianza

Teorema del Límite Central. Inferencia de un parámetro poblacional a partir de un estimador muestral

Descriptive Statistics

	N	Mean	Std. Deviation
Age (years)	236	57,07	14,10
Valid N (listwise)	236		

Descriptive Statistics

	N	Mean	Std. Deviation
Age (years)	238	56,46	13,26
Valid N (listwise)	238		

Descriptive Statistics

	N	Mean	Std. Deviation
Age (years)	243	56,86	13,30
Valid N (listwise)	243		

Descriptive Statistics

	N	Mean	Std. Deviation
Age (years)	258	56,94	13,63
Valid N (listwise)	258		

Descriptive Statistics

	N	Mean	Std. Deviation
Age (years)	1207	56,39	13,33
Valid N (listwise)	1207		

Descriptive Statistics

	N	Mean	Std. Deviation
Age (years)	236	55,00	13,35
Valid N (listwise)	236		

Descriptive Statistics

	N	Mean	Std. Deviation
Age (years)	250	55,83	13,10
Valid N (listwise)	250		

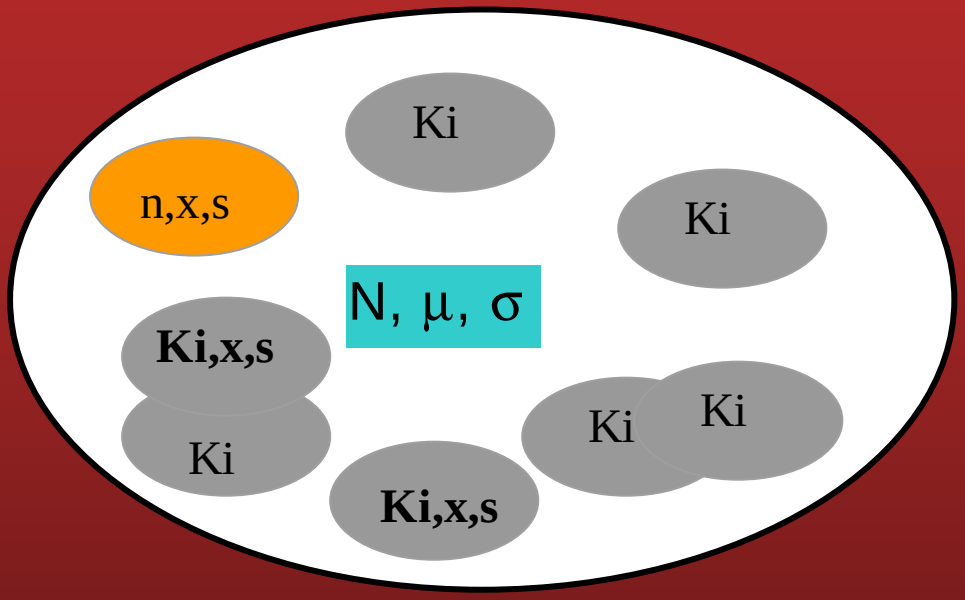
Descriptive Statistics

	N	Mean	Std. Deviation
Age (years)	264	55,94	13,12
Valid N (listwise)	264		

Descriptive Statistics

	N	Mean	Std. Deviation
Age (years)	219	54,93	12,88
Valid N (listwise)	219		

Teorema Central del Límite. Inferencia de un parámetro poblacional a partir de un estimador muestral



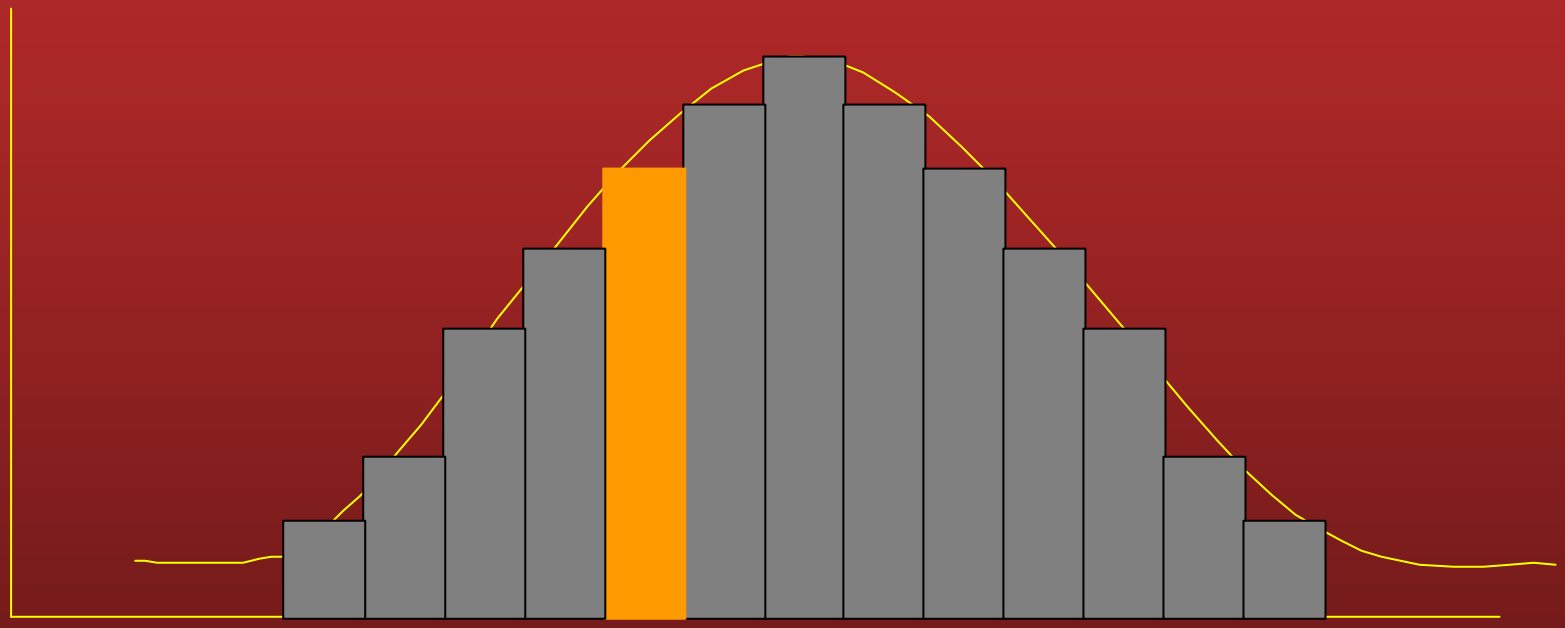
N = tamaño poblacional.
n = tamaño muestral.
k = número de muestras de tamaño n.

$$k = \binom{N}{n} = \frac{N!}{n!(N-n)!}$$

- Sea cual sea la distribución de una **variable aleatoria poblacional**, con media μ y varianza σ^2, la distribución de las medias, x, de muestras de tamaño n es, aproximadamente, una normal, con media μ y varianza σ^2/n(*) (cuanto mayor sea n, menor es la varianza), y tanto más aproximadamente cuanto mayor sea el tamaño de n (por tanto, a mayor n, mayor simetría de la distribución).

(*) La raiz cuadrada de este valor es el "Error estandar del estimador".

Teorema Central del Límite.
Conociendo la media de nuestra muestra podemos inferir la media de la población de donde se obtuvo.



-1.96*σ/√n μ 1.96*σ/√n

-1.96 0 1.96

$x_i \rightarrow N\left(m, s^2/n\right)$

Tipificación

$$\frac{\bar{x}_i - m}{s/\sqrt{n}} \rightarrow N(0,1)$$

Distribución empírica

Distribución estandarizada

Cálculo del Intervalo de Confianza de un estimador

$$\mathbf{P}\left(-1.96 \leq \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq +1.96\right) = 0.95$$

$$P\left(-1.96\frac{s}{\sqrt{n}} \leq \bar{x} - m \leq +1.96\frac{s}{\sqrt{n}}\right) = 0.95$$

$$P\left(\bar{x} - 1.96\frac{s}{\sqrt{n}} \leq m \leq \bar{x} + 1.96\frac{s}{\sqrt{n}}\right) = 0.95$$

Parámetro (Θ) $\in$ Estimador (Θ) ± 1.96 x (error estandar del estimador)

Error estandar y Error asociado a la estimación o error de muestreo.

Parámetro (Θ) $\in$ Estimador (Θ) $\pm$ 1.96 x (error estandar del estimador)

- **El producto $1.96\frac{s}{\sqrt{n}}$ se denomina "ERROR ASOCIADO A LA ESTIMACIÓN ó ERROR DE MUESTREO".**

- **El valor $\frac{s}{\sqrt{n}}$ se denomina ERROR ESTANDAR de la media (del estimador en general), es decir la desviación estandar de las medias muestrales (recuerda que mide la bondad de ajuste).**

Utilidad del Intervalo de Confianza.

- **¿Porqué utilizar un único punto de corte cuando la elección de ese punto es arbitraria?.**

- **¿Porqué reducir la cuestión de si un tratamiento es efectivo a un valor dicotómico (si/no) cuando sería más apropiado verlo como un continuum?.** **Guyatt G, Jaenschke R, Heddle N, et al. Basic statistics for clinicians. 1. Hypothesis testing. Can Med Assoc J 1995;152:27-32.**

- **El IC proporciona un recorrido de posibilidades para el valor poblacional en la escala de magnitud de la variable medida, no una dicotomía arbitraria basada tan solo en la significación estadística.**

Intervalo de confianza para medias y proporciones

- **La fórmula del IC para un estimador de una media es:**

$$P(\bar{x} - t_{1-\alpha/2}\frac{s}{\sqrt{n}} \le m \le \bar{x} + t_{1-\alpha/2}\frac{s}{\sqrt{n}}) = 1-\alpha$$

- **De manera similar, la fórmula del IC para una proporción es:**

$$P(\hat{p} - t_{1-\alpha/2}\frac{\sqrt{\hat{p}\hat{q}}}{\sqrt{n}} \le p \le \hat{p} + t_{1-\alpha/2}\frac{\sqrt{\hat{p}\hat{q}}}{\sqrt{n}}) = 1-\alpha$$

Test de Hipótesis. Objetivo: rechazar la hipótesis nula.

- **Ho º frecuencia de cáncer ó los años de vida son los mismos en los sujetos fumadores que en los no fumadores.**

$$H_0 \equiv p(A) = p(B) \Rightarrow p(A) - p(B) = 0 \qquad H_0 \equiv p(A) = p(B) \Rightarrow \frac{p(A)}{p(B)} = 1$$

$$H_0 \equiv \bar{x}(A) = \bar{x}(B) \Rightarrow \bar{x}(A) - \bar{x}(B) = 0$$

- **La decisión del investigador estará basada en los datos obtenidos en su muestra (información empírica).**
- **Si la probabilidad (que vamos a buscar en la tabla con la distribución correspondiente al final del libro: número de casos y grados de libertad) de obtener unos resultados tan extremos o más que los encontrados en nuestro estudio es demasiado pequeña, nos atreveremos a rechazar H_0.**

Un ejemplo para entender el intervalo de confianza

FIGURA 1. Tensiones arteriales sistólicas en 100 diabéticos y en 100 no diabéticos. Las medias son respectivamente 146.4 y 140,4 mm Hg. La diferencia de 6,0 mm Hg entre las medias muestrales se indica a la derecha junto con el intervalo de confianza de 95% que va de 1,1 a 10,9 mm Hg.

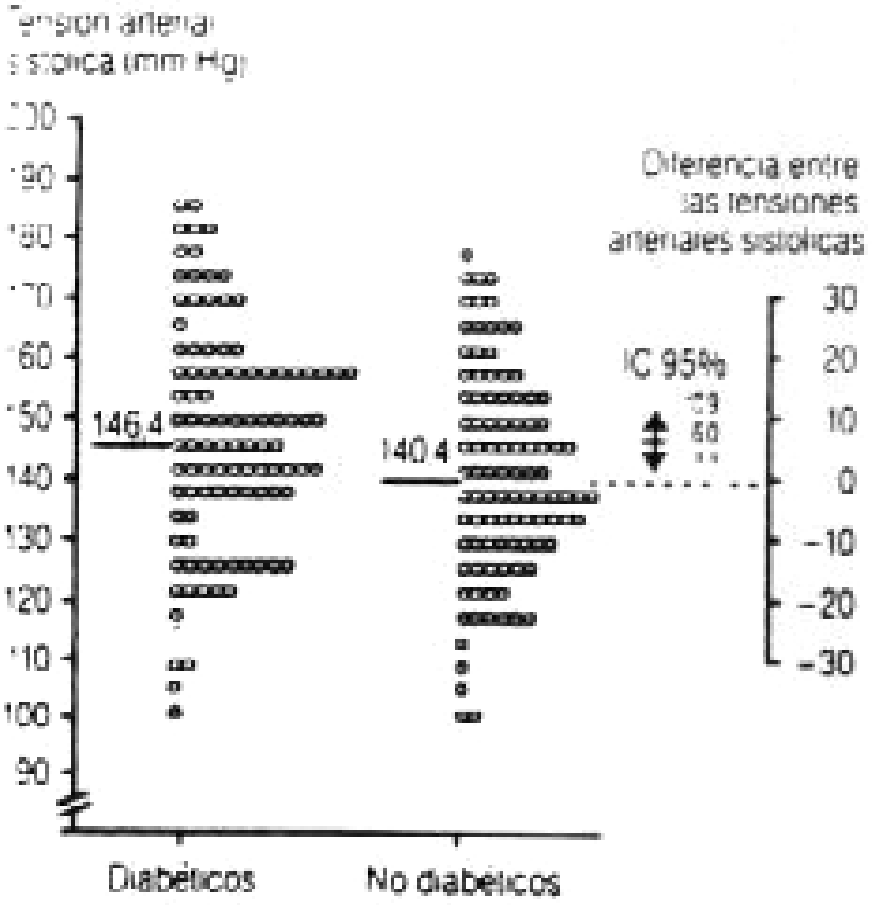
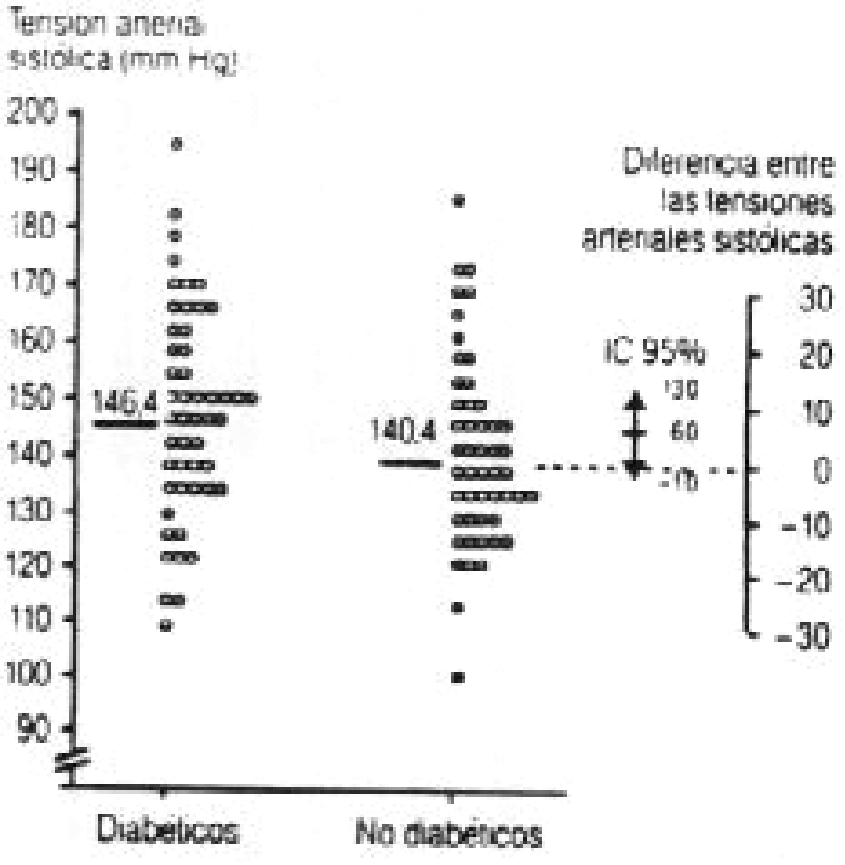


FIGURA 2. Esquema similar al de la figura 1, en este caso con los resultados de dos muestras de tamaño igual a la mitad de las muestras anteriores, es decir, 50 sujetos cada una. Las medias y desviaciones estándar son iguales a las de la figura 1, pero el intervalo de confianza de 95% es más amplio, de −1,0 a 13,0 mm Hg, debido al menor tamaño de las muestras



Intervalo de confianza. Algunas ideas sueltas.

- **La magnitud del estimador indica la importancia del efecto y la amplitud del IC la cantidad de variabilidad (incertidumbre) inherente a la estimaciòn (muestreo).**
- **Cualquier estadístico tiene su error estandar y por tanto su IC.**
- **Si el IC no incluye el valor nulo (para diferencias, 0, y para cocientes, 1), la p< 0,05 por definición.**
- **Cuanto menor sea la muestra, los resultados estarán más expuestos a no indicar la realidad de la población sobre la que se desean inferir resultados, debido a la gran variabilidad del muestreo y al azar.**

¿què sucede si queremos tener màs o menos confianza (99%, 90%) en que el parámetro poblacional se encuentre en nuestro intervalo?

¿Si aumentamos o disminuimos el tamaño muestral?

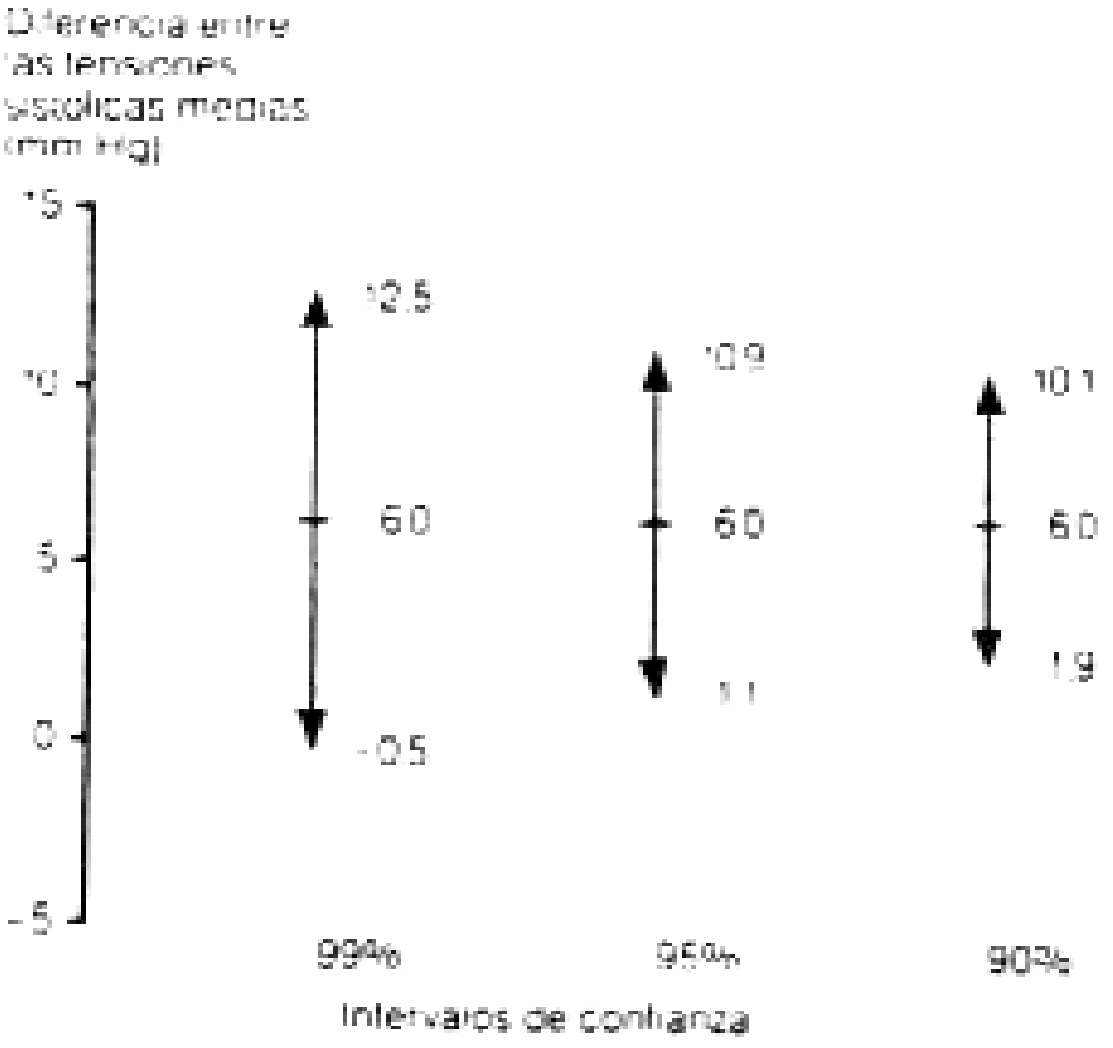
$$P(\bar{x} - 1.96\frac{s}{\sqrt{n}} \leq m \leq \bar{x} + 1.96\frac{s}{\sqrt{n}}) = 0.95$$

Tabla de valores del estadístico Z para α y β

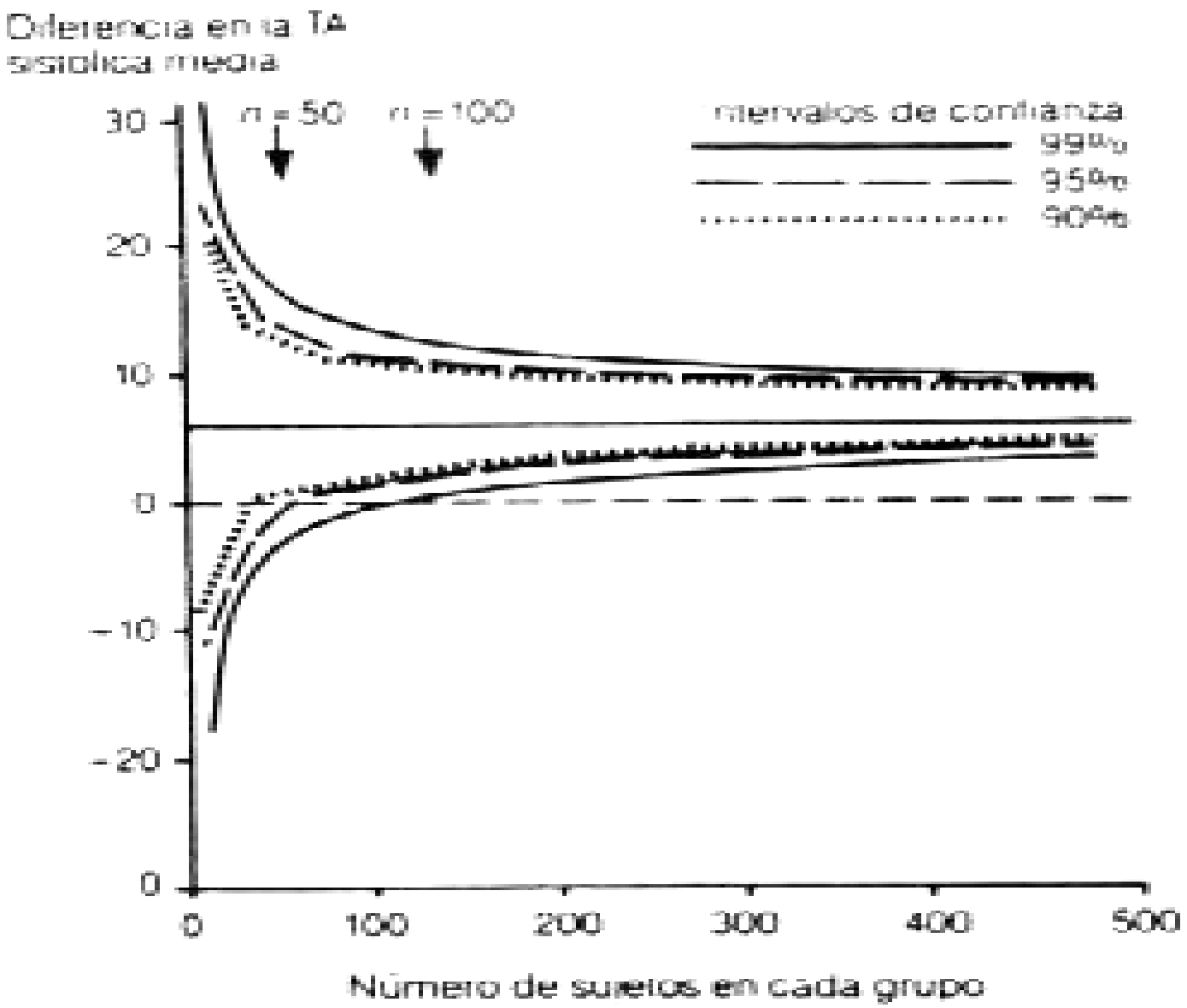
Nivel	Estadístico Z
α 0.90 bilateral	1.64
α 0.95 unilateral	1.64
α 0.95 bilateral	1.96
α 0.99 bilateral	2.57
α 0.999 bilateral	3.29
β 0.20	0.84
β 0.10	1.28
β 0.05	1.64
β 0.01	2.32

Intervalo de Confianza. Cambios con el nivel de confianza

FIGURA 3. Intervalos de confianza asociados con diferentes grados de "confianza", usando los mismos datos de la figura 1

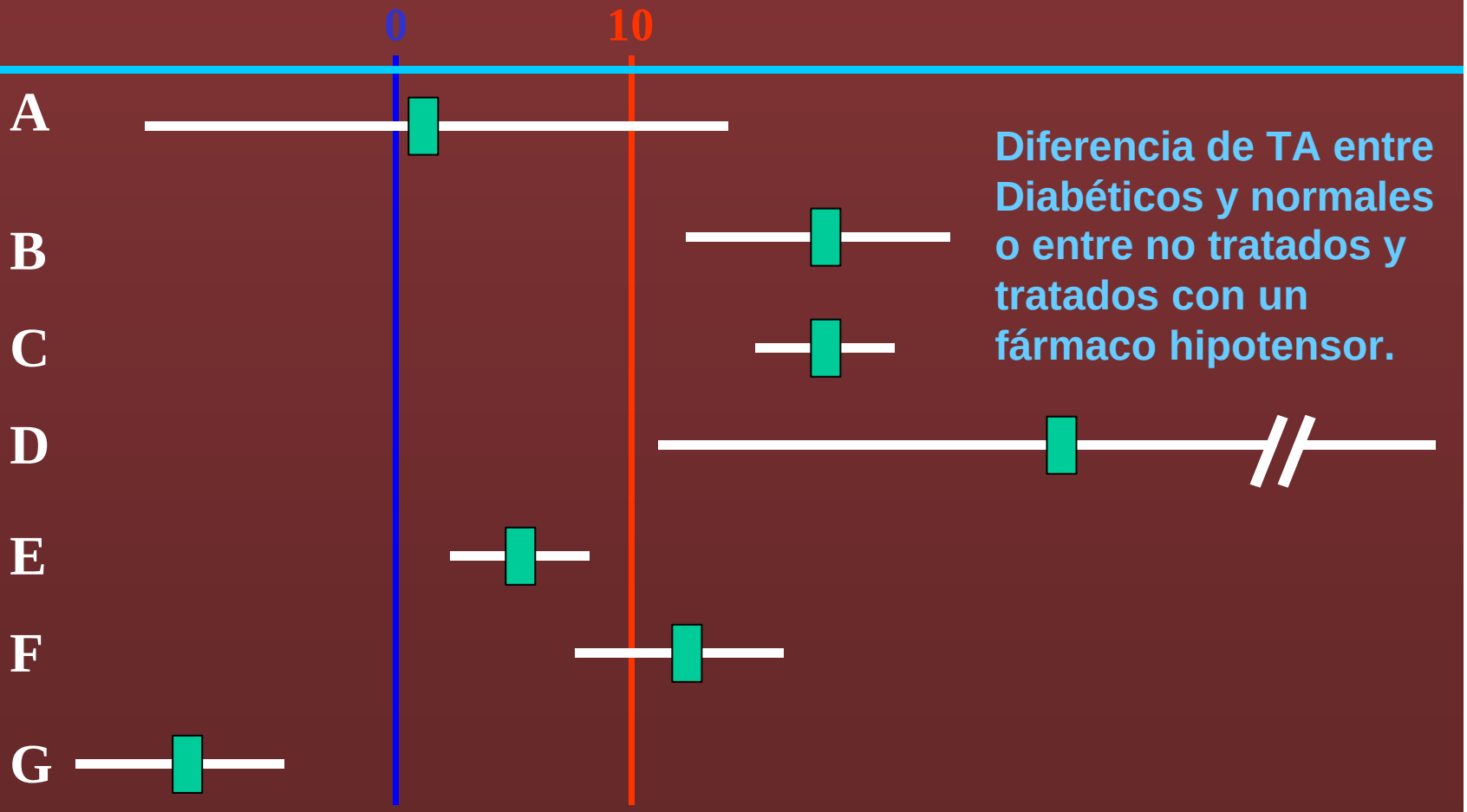


Intervalo de Confianza. Cambios con el nivel de confianza y el tamaño muestral.



Intervalos de Confianza de los estimadores(*)

Decidir lo que constituye una diferencia clínicamente importante es difícil e inevitablemente arbitrario. Las características del resultado que se evalúa (prevenir una muerte, ictus grave, TIA), el riesgo basal, los efectos adversos, los inconvenientes y el costo, todo influye.



Diferencia de TA entre Diabéticos y normales o entre no tratados y tratados con un fármaco hipotensor.

(*) Recuerda, estamos interesados en conocer la magnitud del efecto (cuantía y dirección) y no tanto en saber la probabilidad de que hallamos llegado a un resultado falso positivo (si podemos o no rechazar la hipótesis nula).

Cálculo del tamaño muestral

Cálculo del tamaño muestral. Estimación de una media

$$P(-1.96\frac{s}{\sqrt{n}} \leq \bar{x} - m \leq +1.96\frac{s}{\sqrt{n}}) = 0.95$$

$$P(\bar{x} - m \leq \pm 1.96\frac{s}{\sqrt{n}}) = 0.95$$

La distancia entre x y m no puede ser mayor que el máximo error asociado a la estimación posible para tener una seguridad del 95% de haber elegido una de las k muestras de la población que nos interesa.

$$\bar{x} - m \leq \pm 1.96\frac{s}{\sqrt{n}} \qquad \bar{x} - m < d$$

El producto $1.96\frac{s}{\sqrt{n}}$ se denomina "ERROR ASOCIADO A LA ESTIMACIÓN ó ERROR DE MUESTREO".

El valor $\frac{s}{\sqrt{n}}$ se denomina "ERROR ESTANDAR" de la media (del estimador en general).

Cálculo del tamaño muestral. Estimación de una media

$$\bar{x} - m \leq \pm 1.96 \frac{s}{\sqrt{n}} \qquad \bar{x} - m < d$$

$$d = 1.96 \frac{s}{\sqrt{n}}$$

$$n = (1.96)^2 \frac{s^2}{d^2}$$

Tomando muestras de, como mínimo, el tamaño n, podemos asegurar que en el (1-a)%, usualmente 95%, de ellas obtendremos medias que no diferirán de m en màs de la cantidad prefijada, d.

Cálculo del tamaño muestral. Estimación de una media y una proporción. Factores a considerar

$$n = (1.96)^2 \frac{s^2}{d^2}$$

$$n = (1.96)^2 \frac{pq}{d^2}$$

En consecuencia, el **tamaño muestral dependerá de las pretensiones del investigador** (lo que indica la subjetividad inherente), ya que éste establecerá las pautas de su estudio.

Test de hipótesis

Test de Hipótesis.

- **Test de hipótesis son procedimientos estadísticos que nos permiten decidir acerca de una hipótesis establecida sobre el valor de uno o más parámetros desconocidos (relaciones entre variables); es decir,**

- **Establecer el grado de consistencia entre la hipótesis establecida y la experiencia realizada.**

- **El objetivo de los test de hipótesis consiste en ver si se puede rechazar la hipòtesis de independencia entre dos variables (por ejemplo fumar y cáncer de pulmón). La hipótesis formulada de esta manera se llama Hipótesis nula, Ho.**

Test de Hipótesis. Evaluación de una muestra.

Variable resultado (dependiente):

- **Frecuencia de cáncer de pulmón.**
- **Media de años de supervivencia.**

El objetivo será saber si los valores de las variables obtenidas en nuestra muestra son posibles en la población de referencia.

Test de Hipótesis. Evaluación de dos muestras.

Variable predictora (independiente, factor):

- **Muestra A: Grupo de no fumadores.**
- **Muestra B: Grupo de fumadores.**

Variable resultado (dependiente):

- **Frecuencia de cáncer de pulmón.**
- **Mediana de años de supervivencia.**

Test de Hipótesis. Objetivo: rechazar la hipótesis nula.

- **Ho º frecuencia de cáncer ó los años de vida son los mismos en los sujetos fumadores que en los no fumadores.**

$$H_0 \equiv p(A) = p(B) \Rightarrow p(A) - p(B) = 0 \qquad H_0 \equiv p(A) = p(B) \Rightarrow \frac{p(A)}{p(B)} = 1$$

$$H_0 \equiv \bar{x}(A) = \bar{x}(B) \Rightarrow \bar{x}(A) - \bar{x}(B) = 0$$

- **La decisión del investigador estará basada en los datos obtenidos en su muestra (información empírica).**
- **Si la probabilidad de obtener unos resultados tan extremos o más que los encontrados en nuestro estudio es demasiado pequeña, nos atreveremos a rechazar H_0.**

Test de Hipótesis. Aquí viene la p.

- **Ahora bien, siempre hay alguna probabilidad de que nos equivoquemos, porque la realidad (la población teórica de donde obtuvimos nuestra muestra) nunca estará a nuestro alcance.**

	DECISION INVESTIGADOR	
REALIDAD	No rechazar H_o	Rechazar H_o
H_o verdadera	Correcto	Error **a**, tipo I
H_0 falsa	Error **b**, tipo II	Correcto (1-**b**)

- **El test estadístico que aplicamos nos proporciona la probabilidad (p) de equivocarnos al rechazar H_0.**
- **(1- b): Potencia del test.**

Test de Hipótesis. ¿quién calcula la p(1)?

- **Naturalmente, ahora lo hacen las máquinas.**
  - **Supongamos que conocemos que el verdadero valor del colesterol de una población de varones con IAM es 240 mg/dl y la d.e. es 40 mg/dl,**
  - **¿cuál es la probabilidad de obtener una muestra de 100 sujetos seleccionados aleatoriamente de esa población con una media de 260 mg/dl? dicho de otra manera,**
  - **¿cuál es la probabilidad de que un grupo de 100 sujetos de esa población, muestreados al azar, tenga una media de Colesterol de 260 mg/dl?.**
  - **Estandarizamos y obtenemos la puntuación Z.**

$$Z_i = \frac{(x_i - m_i)}{s/\sqrt{n}} = \frac{\text{valor observado} - \text{valor poblacional}}{\text{error estandar del estimador}} \rightarrow N(0,1)$$

Test de Hipótesis. ¿quién calcula la p (2)?

- **Naturalmente, ahora lo hacen las máquinas.**

$$Z_i = \frac{(x_i - m\ )}{s/\sqrt{n}} = \frac{260 - 240}{40/\sqrt{100}} = 5$$

- **una media de 260 mg/dl se coloca a 5 veces el error estandar de la media poblacional de 240 mg/dl.**
- **La probabilidad de interés será < 0.001.**
- **La probabilidad de que una muestra de 100 sujetos con nivel medio de colesterol de 260 mg/dl pertenezca a la población de referencia es muy pequeña; sería muy raro que esta muestra perteneciera a la población referida.**
- **En consecuencia, nos atrevemos a rechazar la hipótesis nula de que esta muestra de 100 sujetos pertenece a la población de referencia.**

Test de Hipótesis. Aquí viene la p.

- **Debe recordarse siempre que aceptar H_o no significa que sea verdadera.**

- **Para un tamaño de muestras suficientemente grandes siempre podremos demostrar diferencias estadísticamente significativas.**

$$Z_i = \frac{(x_i - m)}{s/\sqrt{n}} = \frac{\text{valor observado} - \text{valor poblacional}}{\text{error estandar del estimador}} \rightarrow N(0,1)$$

- **Es fundamental establecer la diferencia clínicamente relevante más que la estadísticamente significativa.**

Puntos a destacar en la valoración de los aspectos estadísticos de un estudio

- **¿Han elegido los autores el escenario de forma correcta?**
  - **¿Han establecido si los grupos son comparables y si es necesario, han realizado ajustes para las diferencias basales?.**
  - **¿Qué tipo de datos han utilizado?. ¿Han usado los tests estadísticos apropiados?.**
  - **Si los tests estadísticos son oscuros, ¿porqué han decidido utilizarlos?.**
  - **¿Los datos han sido analizados de acuerdo al protocolo original del estudio?.**

Greenhalgh T. How to read a paper. The basics of evidence based medicine. London:BMJ Publishing Group. 1997.

Puntos a destacar en la valoración de los aspectos estadísticos de un estudio

- **Datos apareados, colas y sujetos extremos.**
  - **¿Se han realizado tests apareados con datos apareados?.**
  - **¿Se ha realizado una prueba de dos colas si el efecto de la intervención puede ser también de tipo negativo?.**
  - **¿Se ha tratado el problema de los sujetos extremos con sentido común y con los ajustes estadísticos apropiados?.**

Greenhalgh T. How to read a paper. The basics of evidence based medicine. London:BMJ Publishing Group. 1997.

Puntos a destacar en la valoración de los aspectos estadísticos de un estudio

- **Correlación, regresión y causalidad:**
  - **¿Se ha distinguido correlación de regresión y se ha calculado e interpretado correctamente el coeficiente de correlación?.**
  - **¿Se han realizado asunciones sobre la naturaleza y dirección de la causalidad?**
- **Probabilidad y confianza.**
  - **¿Se han calculado e interpretado correctamente los valores de la p.**
  - **Se han calculado los intervalos de confianza y los reflejan las conclusiones de los autores?.**

Greenhalgh T. How to read a paper. The basics of evidence based medicine. London:BMJ Publishing Group. 1997.

Puntos a destacar en la valoración de los aspectos estadísticos de un estudio

- **¿Se han expresado los resultados en términos de la probabilidad de daño o beneficio que puede esperarse en un paciente individual?**
  - **Riesgo Relativo/Odds ratio.**
  - **Reducción del Riesgo Relativo.**
  - **Reducción del Riesgo Absoluto.**
  - **Número Necesario para Tratar.**

Greenhalgh T. How to read a paper. The basics of evidence based medicine. London:BMJ Publishing Group. 1997.

Tipo de test estadístico a utilizar para hacer inferencias (estimación de parámetros poblacionales o comparación entre muestras).

DISTRIBUCION	VARIABLE INDEPENDIENTE	VARIABLE DEPENDIENTE	RELACIÓN ENTRE LAS MUESTRAS	TEST ESTADÍSTICO
UNIVARIANTE O BIVARIANTE				
Normal (Paramétricos)	Una sola muestra (compara con valor teórico)	Cuantitativa Cualitativa		t-student para una muestra Chi-cuadrado para una muestra
	Dicotómica	Categórica	No relacionadas	No existe (se usa Chi-cuadrado de Pearson)
			Relacionadas	Se usa el Test de McNemar
		Cuantitativa	Relacionadas	t-student muestras dependientes
			No relacionadas	t-student muestras independientes
	Policotómica	Categórica	No relacionadas	No existe (se usa Chi-cuadrado de Pearson
		Cuantitativa	No relacionadas	ANOVA de una vía (ONEWAY) ANOVA de dos vías

Tipo de test estadístico a utilizar para hacer inferencias (estimación de parámetros poblacionales o comparación entre muestras).

DISTRIBUCION	VARIABLE INDEPENDIENTE	VARIABLE DEPENDIENTE	RELACIÓN ENTRE LAS MUESTRAS	TEST ESTADÍSTICO
UNIVARIANTE O BIVARIANTE				
No normal (No paramétri-cos)	Una sola muestra (compara con valor teórico)			Binomial Chi-cuadrado de Pearson Chi-cuadrado de Mantel-Haenzsel Kolmogorow-Smirnov Rachas
	Dicotómica	Categórica	Relacionadas	Test exacto de McNemar Prueba de los Signos
			No relacionadas	Chi-cuadrado de Pearson Test exacto de Fisher
		Cuantitativas	Relacionadas	Prueba de los signos Prueba de los rangos con signo de Wilcoxon
			No relacionadas	Mann-Whitney W de Wilcoxon Mediana Z Kolmogorov-Smirnov Rachas de Wald-Wolfowitz Valores extremos de Moses
	Policotómica	Categórica	No relacionadas	Prueba Q de Cochran
		Cuantitativa	Relacionadas	Friedman W de Kendall (concordancia)
			No relacionadas	Kruskal-Wallis Mediana K variables ANOVA de dos vías por rangos

tras).

DISTRIBUCION	VARIABLE INDEPENDIENTE	VARIABLE DEPENDIENTE	RELACIÓN ENTRE LAS MUESTRAS	TEST ESTADÍSTICO
UNIVARIANTE O BIVARIANTE				
COVARIACION (medidas de dos variables en los mismos sujetos o unidades de análisis del estudio)				
Paramétrico	Ambas variables cuantitativas, pero no hay dependiente o independiente			Correlación de Pearson
No paramétrico	Ambas variables cuantitativas, pero no hay dependiente o independiente			Correlación de Spearman Correlación Tau de Kendal
REGRESIÓN				
Paramétrica	Cuantitativa o Cualitativa	Cuantitativa		Regresión lineal simple

Asunciones para el uso de test paramétricos. Distribución normal

- **Distribución normal de la variable dependiente para los dos niveles del factor:**
- **Pruebas estadísticas:**
  - **Test de Kolmogorov-Smirnov con la correccion de Lilliefors y test de Shapiro-Wilk.**
  - **Asimetría y curtosis.**
- **Evaluación gráfica:**
  - **Stem & Leaf.**
  - **normal probability plot (Q-Q plot).**
  - **detrended normal plot (P-P plot).**
  - **Histograma y curva normal: Permite echar un ojo al aspecto del histograma valorando su distribución.**

Asunciones para el uso de test paramétricos: Homogeneidad de las varianzas.

- **Homogeneidad de varianzas de la variable dependiente para cada nivel del factor que se analiza en el modelo.**
- **Pruebas estadísticas:**
  - **Test de Levene.**
  - **Asimetría (uno u otro lado) y curtosis (alejadas o próximas) .**
- **Representación gráfica**
  - **Spread vs. Level**
  - **Scatter plots entre observados, predichos y residuales estandarizados.**

Asunciones para el uso de test paramétricos: Variable medida al menos en escala de intervalo.

- **La variable dependiente debe medirse al menos en una escala de intervalo, de forma que existe la misma proporción entre dos valores consecutivos de la escala.**

Asunciones para el uso de test paramétricos: Independencia de las observaciones.

- **Independencia de las observaciones con distribución aleatoria de la muestra en los diferentes subgrupos formados por las combinaciones de niveles de los factores. Esto significa que la puntuación obtenida por un sujeto es independiente de la que obtiene otro.**
- **Prueba estadística:**
  - **Test de las rachas.**

Pruebas de inferencia estadística. Comparación de medias entre dos grupos. T de student y pruebas no paramétricas.

- T de studen para una muestra.
- T de student para muestras independientes.
- T de student para muestras dependientes.
- Pruebas no paramétricas:
  - Test de Mann-Whitney
  - Test de Wilcoxon

Pruebas de inferencia estadística.
Comparación de proporciones entre dos o mas grupos.
Test estadístico chi-cuadrado.

- **Test chi cuadrado:**
  - **Pearson.**
  - **Corrección de continuidad de Yates.**
  - **Razón de verosimilitud.**
  - **Prueba exacta de Fisher.**
  - **Asociación lineal.**
- **Prueba de McNemar**
- **Prueba Q de Cochran.**

Pruebas de inferencia estadística.
Comparación de medias entre dos o mas grupos.
Modelos Lineales Generalizados.
ANOVA y pruebas no paramétricas.

- **ANOVA de una vía (factor).**
- **ANOVA de dos o más vías (factores).**
- **ANCOVA.**
- **MANOVA (análisis multivariante de la varianza).**
- **ANOVA de medidas repetidas.**